

On Nonrepetitive Sequences

R. C. ENTRINGER AND D. E. JACKSON

University of New Mexico, Albuquerque, New Mexico, 87106

AND

J. A. SCHATZ*

Sandia Corporation, Albuquerque, New Mexico, 87115

Communicated by P. Erdős

Received May 8, 1972

We prove the following results: (1) There exists an infinite binary sequence having no identical adjacent blocks of length 3 or greater. (2) Every binary sequence of length greater than 18 has identical adjacent blocks of length 2 or greater. (3) Every infinite binary sequence has arbitrarily long adjacent blocks that are permutations of each other.

1. Let X be a finite set (alphabet), and let $S = (x_1, x_2, x_3, \dots)$ be a finite or infinite sequence of elements of X . By a block B of S , we shall mean a finite subsequence $(x_{i+1}, x_{i+2}, \dots, x_{i+k})$ for some integers $i \geq 0$ and $k \geq 1$; we shall call k the block length. In 1906, Thue [5] showed that (1) when X has 3 symbols there is an infinite S which has no two adjacent blocks which are identical, and (2) when X has 2 symbols there is an infinite S which has no 3 adjacent blocks which are identical. (Any sequence of length at least 4 on two symbols has two consecutive occurrences of the same block.) These results have been repeatedly rediscovered over the years and have been used in many different applications. Hedlund [3] lists some of these rediscoveries and discusses applications of these results. Brown [1] lists some later rediscoveries.

In this paper, we consider only the case $X = \{0, 1\}$. Despite the repeated rediscovery of result (2), little consideration seems to have been given to asking whether we can find an infinite S with restrictions on the blocks B for which there are adjacent occurrences of B in S . In this connection, we consider the following conjecture.

* This work was supported by the United States Atomic Energy Commission.

CONJECTURE A. *Every infinite sequence S on 2 symbols has arbitrarily long identical adjacent blocks.*

The only statement which we have found seems to say that Conjecture A is true [2, page 240]. In Section 2, we show that this is false. Theorem 1 shows that we can choose S so that the longest repeated block has length 2, and Theorem 2 shows that this is the best possible result. At the end of Section 2 we return to Conjecture A and state an unresolved question.

Section 3 considers the status of Conjecture A if we are interested in strongly non-repetitive sequences. A sequence is strongly non-repetitive if it contains no two adjacent blocks which are permutations of each other. The terminology is that of Pleasants [4]. The status of the problems for alphabets of more than two symbols is given by Brown [1]. Theorem 3 shows that any infinite sequence on two symbols has arbitrarily long blocks which are permutations of each other.

2. Theorem 1 shows that Conjecture A is false.

THEOREM 1. *There exists an infinite binary sequence having no identical adjacent blocks of length 3 or greater.*

Proof. Choose any infinite sequence S on three symbols a , b , and c which has no identical adjacent blocks. In this sequence replace a by 1010, b by 1100, and c by 0111 to form a new sequence $B = (b_1, b_2, \dots)$ which we shall prove has the property stated in the theorem.

By *word* we will mean one of the quadruples 1010, 1100, or 0111 (we suppress commas occasionally).

We will make repeated use of the following properties of our three words. Each property is easily verified.

PROPERTY P. The block $(b_{n+1}, b_{n+2}, b_{n+3}, b_{n+4})$ is a word if and only if n is a multiple of 4, i.e., a word cannot be formed by using the last bits of a word w and the first bits of a word distinct from w .

PROPERTY Q. Each word is determined by its first two bits and also by its last two bits.

PROPERTY R. No word has the same first and last bits.

The remainder of the proof is divided into three parts:

(i) We first show that, if B has identical adjacent blocks of length $k \geq 4$, then k is a multiple of 4.

Suppose $(b_{n+1}, \dots, b_{n+k})$ and $(b_{n+k+1}, \dots, b_{n+2k})$ are identical adjacent blocks of B . We may assume $k \geq 5$. There are two cases to consider. Either one of the blocks (and therefore the other) contains an entire word, or $k = 5$ and $(b_{n+4}, b_{n+5}, b_{n+6}, b_{n+7})$ is a word.

In the first case we may assume $(b_{n+i}, b_{n+i+1}, b_{n+i+2}, b_{n+i+3})$ is a word for some i , $1 \leq i \leq k-3$, so that $(b_{n+k+i}, b_{n+k+i+1}, b_{n+k+i+2}, b_{n+k+i+3})$ is also a word. But then by Property P we have $4 \mid (n+i+3)$ and $4 \mid (n+k+i+3)$, so that $4 \mid k$.

In the second case we have $b_{n+1} = b_{n+6}$ and $b_{n+2} = b_{n+7}$ so that (b_{n+6}, b_{n+7}) constitutes both the last pair of bits of some word w and the middle pair of bits of some word distinct from w . This can occur only if $(b_{n+6}, b_{n+7}) = (1, 0)$, and this in turn only if $(b_{n+1}, b_{n+2}) = (1, 0) = (b_{n+4}, b_{n+5}) = (b_{n+9}, b_{n+10})$. As a consequence we have $b_{n+3} = 0$ and $b_{n+8} = 1$, which cannot happen since the blocks were assumed to be identical.

(ii) We next show that, if B has identical adjacent blocks of length $k \geq 4$, then the original sequence S had identical adjacent blocks (contrary to assumption).

In view of (i) and Property P this assertion will follow if we can demonstrate the existence of two identical adjacent blocks $(b_{4m+1}, \dots, b_{4m+k})$ and $(b_{4m+k+1}, \dots, b_{4m+2k})$ whenever we have $(b_{n+1}, \dots, b_{n+k}) = (b_{n+k+1}, \dots, b_{n+2k})$ for some n .

If $n = 4m + 1$ or $4m + 2$ then by Property Q we must have $(b_{4m+1}, b_{4m+2}) = (b_{4m+k+1}, b_{4m+k+2})$ so that

$$(b_{4m+1}, \dots, b_{4m+k}) = (b_{4m+k+1}, \dots, b_{4m+2k}).$$

If $n = 4m - 1$ then by Property Q we have $b_{n+k+1} = b_{n+2k+1}$ so that $(b_{4m+1}, \dots, b_{4m+k}) = (b_{4m+k+1}, \dots, b_{4m+2k})$. This completes the proof of (ii).

(iii) It remains to prove that B has no identical adjacent blocks of length 3.

Assume, to the contrary, that $(b_{n+1}, b_{n+2}, b_{n+3}) = (b_{n+4}, b_{n+5}, b_{n+6})$. Now any word beginning at b_{n+1} , b_{n+2} , or b_{n+3} would end at b_{n+4} , b_{n+5} , or b_{n+6} , respectively. This is impossible in view of Property R. Hence $(b_{n+4}, b_{n+5}, b_{n+6})$ is the beginning of a word and (since it is the same as $(b_{n+1}, b_{n+2}, b_{n+3})$) also the end of a word. Since no words have this property the proof of the theorem is complete.

THEOREM 2. *Every binary sequence of length greater than 18 has identical adjacent blocks of length 2 or greater.*

Proof. The proof is given diagrammatically in Figure 1. The 21

Theorem 1 has shown that Conjecture A is false. The sequence constructed in the proof of Theorem 1 has quintuply repeated adjacent blocks (namely, the block 1). We do not know whether this is necessary. It seems likely that there is no infinite sequence on two symbols which satisfies both result (2) and the negation of Conjecture A. We formalize this in the following conjecture:

CONJECTURE B. *Any infinite sequence on two symbols which has no three adjacent identical blocks has arbitrarily long adjacent identical blocks.*

3. This section is devoted to showing that Conjecture A is true if we ask only that adjacent blocks be permutations of each other.

THEOREM 3. *Every infinite binary sequence has arbitrarily long adjacent blocks that are permutations of each other.*

Proof. We will actually prove the stronger theorem that all finite binary sequences of length $k^2 + 6k$ or greater contain adjacent blocks of length k or greater that are permutations of each other.

Let $B = (b_1, \dots, b_{k^2+6k})$ be any binary sequence of length $k^2 + 6k$ and set $f(m)$, $0 \leq m \leq k^2 + 5k$, equal to the number of 1's in the block $(b_{m+1}, \dots, b_{m+k})$. Also set $g(n) = f(nk)$ for $0 \leq n \leq k + 5$. Then since $0 \leq g(n) \leq k$ there is a p , $3 \leq p \leq k + 2$, such that either $g(p) \geq g(p-1)$ and $g(p) \geq g(p+1)$, or else $g(p) \leq g(p-1)$ and $g(p) \leq g(p+1)$. Replacing B by its complement if necessary, we assume that $g(p) \geq g(p-1)$ and $g(p) \geq g(p+1)$.

Now the function $h(i) = f(pk+i) - f((p-1)k+i)$ satisfies $h(0) \geq 0$, $h(k) \leq 0$, and $|h(i+1) - h(i)| \leq 2$ (since $|f(m+1) - f(m)| \leq 1$) so that either there is a j for which $h(j) = 0$ and we are done or for some j , $0 \leq j \leq k$, we have $h(j) = 1$ and $h(j+1) = -1$.

In the latter case B has adjacent blocks $(b_{(p-1)k+j+1}, \dots, b_{pk+j})$ and $(b_{pk+j+1}, \dots, b_{(p+1)k+j})$ having $f(pk+j) - 1$ and $f(pk+j)$ 1's respectively. Furthermore $h(j+1) = -1$ implies $b_{pk+j+1} = 1$ and $b_{(p+1)k+j+1} = 0$.

Now let r be the largest integer less than $(p-1)k+j+1$ for which $b_r = 1$. If there is none we are done since $(p-1)k+j \geq 2k$. Let s be the least integer greater than $(p+1)k+j$ for which $b_s = 1$. Again if there is none we are done since $k^2 + 6k - (p+1)k - j \geq 2k$. We note that $s > (p+1)k+j+1$.

If $(p-1)k+j+1-r < s-(p+1)k-j$ then (b_r, \dots, b_{pk+j}) and $(b_{pk+j+1}, \dots, b_{2pk+2j+1-r})$ are adjacent blocks of length $pk+j-r+1 \geq k$ and having $f(pk+j)$ 1's each so that we are done. If

$$(p-1)k+j+1-r \geq s-(p+1)k-j$$

then $(b_{2pk+2j+3-s}, \dots, b_{pk+j+1})$ and (b_{pk+j+2}, \dots, b_s) are adjacent blocks of length $s - pk - j - 1 \geq k$ and having $f(pk + j)$ 1's each. This completes the proof.

REFERENCES

1. T. C. BROWN, Is there a sequence on four symbols in which no two adjacent segments are permutations of one another? *Amer. Math. Monthly* **78** (1971), 886–888.
2. P. ERDÖS, Some unsolved problems, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 221–254.
3. G. A. HEDLUND, Remarks on the work of Axel Thue on sequences, *Nordisk Mat. Tidskr.* **15** (1967), 147–150.
4. P. A. B. PLEASANTS, Non-repetitive sequences, *Proc. Cambridge Philos. Soc.* **68** (1970), 267–274.
5. A. THUE, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr., I. Mat.-Nat. Kl. (Christiania)* No. 7 (1906), 1–22.